**Comparative Genome Assembly**

**Abstract**

Until now, the most complex and computationally intensive tasks of genome sequence analysis is genome assembly we have the resources in both software and hardware that can help us. as we have a great number of sequencing genomes that increase organisms for two or more closely related spices that have been sequenced, that lead to comparative genome assembly

**Introduction**

The most used sequenced genomes as a reference are whole genome shotgun (WGS), which breaks the genome into small fragments and sequence them from both ends to reconstruct the chromosomes of the target organism using latest sequencing technologies. assemblers depended on not only to reconstruct the genome but also to answer basic question. Scientific community has recognized the value of sequencing for example sequencing the genome of (mouse , rat, chimpanzee)and mapping to human to better understand . The algorithm used is overlap layout consensus. Most assemblers use hashing strategies in order to identify those reads that are likely to overlap, another assemblers use (AMOS-Cmp) in this case we skip the overlap, step , reads are aligned to the reference genome using modified version called (MUMmer algorithm) ( alignment layout consensus) .

**The goal to be obtained an assembly of this genome using a reference genome as a template**

**Related Works:**

1. A new system for aligning whole genome sequences is described. Using an efficient data structure called a suffix tree, the system is able to rapidly align sequences containing millions of nucleotides. Its use is demonstrated on two strains of Mycobacterium tuberculosis, on two less similar species of Mycoplasma bacteria and on two syntenic sequences from human chromosome 12 and mouse chromosome 6. In each case it found an alignment of the input sequences, using between 30 s and 2 min of computation time. From the system output, information on single nucleotide changes, translocations and homologous genes can easily be extracted. Use of the algorithm should facilitate analysis of syntenic chromosomal regions, strain-to-strain comparisons, evolutionary comparisons and genomic duplications. ( . Delcher, A. L., Kasif, S., Fleischmann, R. D. et al. (1999), ‘Alignment of whole genomes’, Nucleic Acids Res., Vol. 27(11), pp. 2369–2376.
2. We report on the quality of a whole-genome assembly of *Drosophila melanogaster* and the nature of the computer algorithms that accomplished it. Three independent external data sources essentially agree with and support the assembly's sequence and ordering of contigs across the euchromatic portion of the genome. In addition, there are isolated contigs that we believe represent nonrepetitive pockets within the heterochromatin of the centromeres. Comparison with a previously sequenced 2.9- megabase, region indicates that sequencing accuracy within nonrepetitive segments is greater than 99.99% without manual curation. As such, this initial reconstruction of the *Drosophila* sequence should be of substantial value of the  scientific community.( Myers, E. W., Sutton, G. G., Delcher, A. L. (2000), ‘A whole-genome assembly of Drosophila’, Science, Vol. 287(5461), pp. 2196–2204)